DOCUMENT RESUME

ED 295 987 TM 011 831

AUTHOR Neel, John H.
TITLE Methods for Determining Minimal Effect Sizes in Power and Sample Size Calculations.
PUB DATE Mar 88
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Effect Size; Error of Measurement; Meta Analysis; *Sample Size; Scores; Statistical Analysis
IDENTIFIERS *Minimal Effect Size; *Power (Statistics)

ABSTRACT

Methods to aid the statistical researchers in making arguments as to what constitutes minimal effect size (MES) are outlined. By selecting MES and determining power for that effect, the researcher communicates to the reader what the researcher considers to be important and what the researcher's chances are for finding that effect or a larger one. Methods addressed include: (1) deriving MES from theory in an absolute manner; (2) relying on previous research; (3) assigning a monetary difference to the difference in hypothetical and real values; (4) appealing to authority; (5) accessing very large samples; (6) relying on traditionally accepted minimal sample sizes; (7) relating MES to a variable other that of immediate interest; (8) using grade equivalent scores; (9) using the MES from a meta-analysis; (10) adopting standard error of measurement; and (11) using small, medium, and large effect sizes for determination of power. Three graphs are presented. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295987

METHODS FOR DETERMINING MINIMAL EFFECT SIZES IN POWER AND SAMPLE SIZE CALCULATIONS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JOHN H. NEEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

John H. Neel

Georgia State University

Presented at the American Educational Research Association Convention

New Orleans, Louisiana
March, 1988

TM 011 831

Introduction

Fundamental concepts of the power of statistical hypothesis testing are covered in most standard introductory statistics texts (e.g., Hayes 1973 pp. 357-362, Glass and Stanley, 1970 pp. 283-288). For the sake of a concrete example and for making definitions clear consider the following hypothesis test:

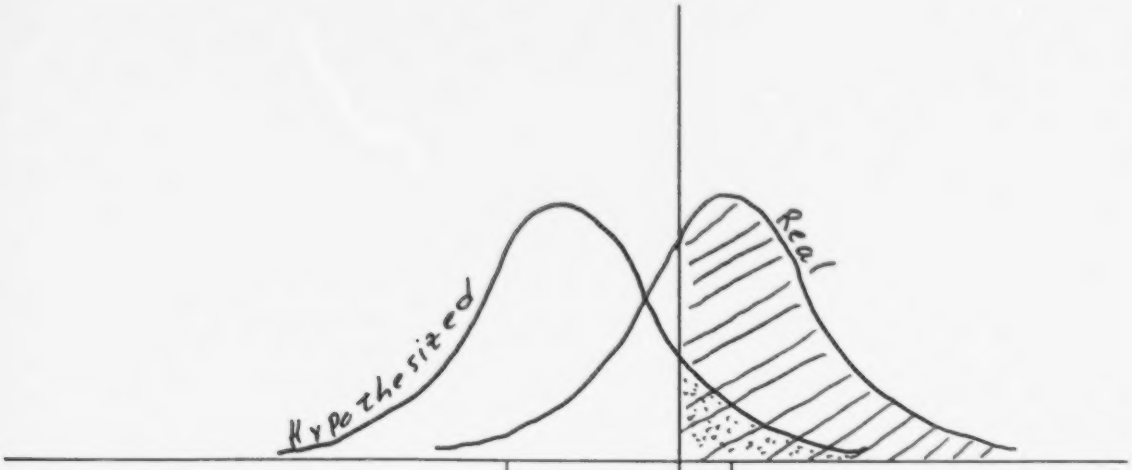
> A researcher believes that a specific diet will result in an increased I.Q. for young children. Nine children have been subjected to the diet and given an I.Q. test after an appropriate length of time. The researcher now wishes to test the hypothesis that the population the nine children represent has a mean I.Q. of 100. He assumes the variance of this population is 225 and thus conducts a z test for one mean (using $\alpha = .05$). This may be summarized as follows:
>
> Ho:μ=100
> Ha:μ=100
> α= .05
> test statistics: z (σ_2 = 225)
> reject Ho if z>1.645

Let us assume that the diet is effective in raising I.Q. and in fact raises the mean population I.Q. to 110; then the relationships between the hypothesized and real means,the hypothesized and real distribution of the sample means, and the critical value of the test statistic are shown in Figure 1. Two distributions

FIGURE 1

Power Calculation for a One Tailed Z test

(sample size = 9, effect size = 10)



$\bar{X}$ $\mu_o = 100$ $\mu_R = 110$

$\bar{X} = 108.25$

z_o $z_o = 1.645$

z_R $z_R = -.35$

μ_o = hypothesized population mean

z_o = critical value of z to yield α of .05

μ_R = real value of population mean

z_R = number of standard deviations of X from μ_R to x

⁘ -- α = .05

/// -- power = .6368

$\bar{X}$ = sample mean corresponding to z_o

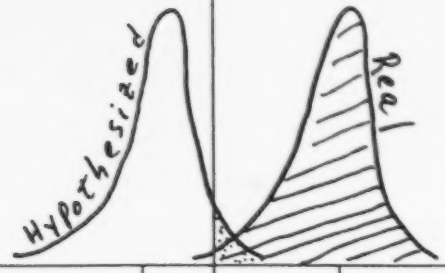
of X are shown in Figure 1, the hypothesized distribution and the real distribution. In addition to the X scale two additional scales are shown; Z_o is the number of standard deviations of X from μ_o to x and Z_R is the number of standard deviations of X from μ_R to x. The decision rule is to reject Ho if Z_o is greater than 1.645 or, equivalently, if X is greater than 108.25; the probability of rejecting Ho is the probability of finding X greater than 108.25. Since the real distribution of X is the distribution with mean equal to 110, the probability of finding X greater than 108.25 is found from the real distribution and using z_R; that probability is found to be .6368. Thus the power of this test given μ_o=100 and μ_R=110 is .6368. The difference between μ_o and μ_R will be termed the effect size in this paper. In general the effect size is the difference between the hypothesized and real value of a parameter.

Once the researcher sorts out the relationships between effect size, statistics, parameters, hypothesized distributions, and real distributions his interest generally focuses on the power and sample size. Textbooks are helpful there and generally give or imply two methods for manipulating the power of a test. Generally, power may be increased by increasing sample size or by looking for a larger effect size; power is generally decreased by decreasing sample size or by looking for a smaller effect size. Increasing power by increasing sample size decreases the variance of the hypothesized and real distributions of the mean, this may be seen by comparing Figure 1 with Figure 2 for an increase from

FIGURE 2

Power Calculation for a One Tailed Z test

(sample size = 25, effect size = 10)



$\bar{X}$ $\mu_o = 100$ $\mu_R = 110$

$\bar{X} = 103.95$

z_o $z_o = 1.645$

z_R $z_R = -2.01$

μ_o = hypothesized population mean

z_o = critical value of z to yield α of .05

μ_R = real value of population mean

z_R = number of standard deviations of X from μ_R to x

⁘ -- α = .05

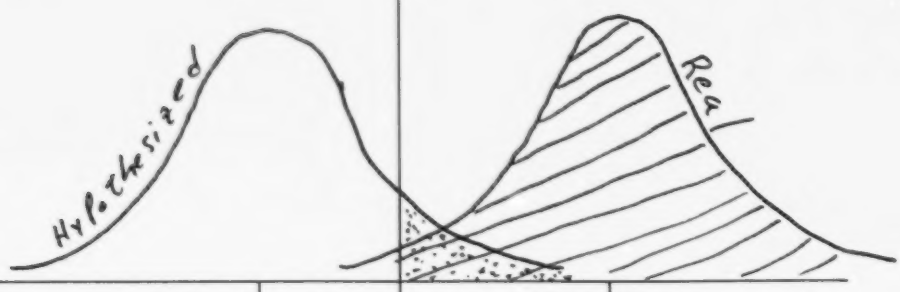
/// -- power = .9778

$\bar{X}$ = sample mean corresponding to z_o

FIGURE 3

Power Calculation for a One Tailed Z test

(sample size = 9, effect size = 20)



$\bar{X}$ $\mu_o = 100$ $\mu_R = 120$

$\bar{X} = 108.25$

z_o $z_o = 1.645$

z_R $z_R = -2.35$

μ_o = hypothesized population mean

z_o = critical value of z to yield α of .05

μ_R = real value of population mean

z_R = number of standard deviations of X from μ_R to x

::: -- α = .05

/// -- power = .9906

$\bar{X}$ = sample mean corresponding to z_o

9 to 25.

Increasing power by increasing the effect size widens the gap between the hypothesized and real distributions, this may be seen by comparing Figure 1 to Figure 3 for an increase from 10 to 20.

For a fixed sample size, power is greater for larger effect sizes and smaller for smaller effect sizes. There is an infinite number of effect sizes and if a researcher can specify a minimum effect size, MES, of interest he can calculate power for that effect size and know that the power for any larger effect size is larger than that for the MES. It is possible then to determine a sample size for the MES and know that sample size will be adequate to insure desired power for the MES and for any larger effect size. The trick is select a reasonable MES.

MES is thus defined as the smallest effect which the researcher would be interested in detecting with a statistical hypothesis test. Similar definitions have been given by Neel (1974), Friedman (1982), Koele (1982), and by Robey and Barcikowski (1984). Kraemer (1985) discusses a "critical effect size" defining it as

> "... a measure expressed in terms of the proposed design and the population characteristics of how strong the theory must minimally be in order to be of interest to society."

Indeed, it is difficult to imagine any other way of dealing with power or sample size other than to find a "worst case" and to find power or sample size for that "worst case". The worst case

here is the case which will require the largest sample size. Each of these authors has defined the situation by somehow referring to the worst case in one way or another as undoubtedly have numerous other authors and instructors in statistics.

It is usually at this point that the statistician puts the burden on the researcher by saying something to the effect that if you will tell me what MES you are interested in, I can tell you what sample size to take to obtain whatever power you desire. Once the MES is determined,the question is then easy to answer with standard techniques [Cohen, 1977) gives methods for calculating power for many statistics covered in a first or second statistics course].

In the diet and I.Q. example given, it seems apparent to the writer that the gain of 30 points in I.Q. is of interest and that a gain of .01 I.Q. points is not of interest. Given some outrageous limits, it is easy to agree that the minimal effect size lies somewhere within those limits. The question then becomes what point between .01 and 10 is the MES?

What constitutes the MES? Is it possible to find that exact point which divides the possible effects into those which are meaningful and those of no import? The answers to these questions are not immediately satisfying. When asked, "What is the smallest effect you are interested in detecting with these hypothesis test?", most researchers the author encounters have no answer. In the discussion with these researchers, two points

become evident. First, the researcher has usually not considered the question previously. Second, the researcher has no idea of how to begin to answer the question. The researcher needs some methods for considering the question. The difficulty is, of course, to define what is meant by the MES in such a manner that other researchers will agree or at least respect and understand why the particular MES was selected.

The question of what constitutes the MES can be answered by presenting an argument for the MES. The researcher can say, "I think the MES is __ because __." The process for the researcher thus becomes one of finding an argument for defending an effect size which is to be the MES. The determination of the MES is dependent on the merit of the argument one makes in favor of the MES. What a researcher needs is a strong argument for the particular MES selected in a study. The list of methods for determining a MES is thus a list of arguments which one might use to defend the selection of a particular MES.

Fourteen methods are presented below for determining a minimal effect size. Each method becomes the argument for the resulting MES. The remainder of this paper is divided into two parts: the first lists and gives examples of the various methods which the writer has found useful in determining MES, the second discusses what power is desirable.

Determining Minimal Effect Size

The following methods for determining minimum effect size

(MES) are listed in what the writer feels is roughly the order of their desirability. The reader may want to order the methods differently, but a different order would not change the utility of determining MES by one of these methods. The order presented here could be followed until the reader establishes a different order of preference for the methods.

1. Absolutely -- As Cohen (1969, p. 12) points out the researcher may expect from theory that a particular effect size exists. In that case he may use the expected effect size as an MES and determine sample size accordingly. Although this is a desirable position for the researcher to find himself in it does not occur frequently in behavioral sciences. Still, if the researcher should find himself in this position he would be on very defensible grounds by finding a sample size which would yield the desired power or detecting the theoretical effect size.

Example: Jensen (1969) claims the correlation between observed I.Q.scores of identical twins is .70. If a researcher felt this were not true, but the true correlation was 0.0 then in choosing a sample size he could use a MES of .70 (.70-0.0). This would insure adequate power to detect a correlation of .70 or higher, if such a correlation existed.

2. Previous Research -- When prior research has been done in an area the researcher may use the results of that research to

obtain an estimate of the minimal effect size which may be of interest to him. The argument here is that in order for the effects of this independent variable to be useful, they must be at least as large as another known effect.

Example: If a previous study had shown a certain reinforcer to yield a mean gain on achievement test of 4.8 points then research involving a different reinforcer would want to detect a mean gain of 4.8 points as well.

3. Monetarily -- In some instances it may be possible to assign a monetary difference to the difference in hypothesized and real values of the hypothesis test and then to determine the MES based on this difference.

Example: If a school system were purchasing hardback dictionaries at $8.00 each, paperback dictionaries at $2.00 each would have to have a life expectancy of at least one-fourth the hardback life expectancy for purchase costs to be equal. Thus if currently purchased hardbacks were known to last 4 years the paperbacks would have to last 1 year to be as cheap in purchase costs. Any statistical testing on the life expectancy of the paperbacks should therefore be concerned with detecting a mean life expectancy of 1 year. (This

example ignores the costs of ordering, distributing, storing, etc.)

4. Appeal To Authority -- When dealing with a test or instrument which has been well researched by some authority the researcher may request that person to estimate the MES. Obviously this makes the sample size determination dependent on the judgement of that authority. As in all cases where judgement is used the researcher and those who read the research should have confidence in the judge.

Example: A student of the writer was working with a well known test and to determine MES it was suggested that the author of the test be contacted and asked his opinion of what difference between the experimental and control groups would be of practical importance. (This was a great success story for the authority felt that a difference of 5 between the means would be of practical importance. This was used as the MES to determine power. A relatively small sample of 25 in experimental and 25 in control group was found necessary to yield a power of .95 with a significance level of .05. This sample was obtained and the resulting observed difference was 5.14 which was significant at $\alpha = .05$).

5. Overkill -- Some researchers are in the fortunate position of having very large samples readily available or even forced upon them. These are generally researchers working in large school systems or projects. If this is the case and the sample will yield a satisfactory power for an effect size which the researcher knows or feels to be smaller than any effect size of practical importance the sample size will be adequate for the MES, whatever that may be.

Example: A Title I Project included 487 students who were given an experimental curriculum. A suitable control group of 300 students was available to test the hypothesis that the mean achievement scores of the population were equal. It was found that this sample size would yield a power of .95 for an effect size of .5 (raw score on the test). Since this was regarded as a very small difference between groups that sample size was regarded as satisfactory.

6. Tradition -- In some areas of research minimal sample sizes are agreed upon by tradition. If one knows of such a tradition and is unable to justify a minimal effect size or sample size by one of the above arguments it would probably be well to abide by tradition and use the minimal size (if one can justify a sample size by one of the above techniques which is smaller than tradi-

tional, he had probably best be prepared to argue his case when he reports to his colleagues.)

Example: In some areas of agricultural research when dealing with correlations the minimal sample size is regarded as 100.

7. Don't -- Sometimes the largest possible sample will be taken in any case and determining MES would not affect the sample size. In this case the researcher should not worry about the minimum effect size but after the data have been gathered should investigate what effect sizes he had some reasonable power for detecting. This is especially important if the power were so small that effect sizes which were of importance yielded low power. If this were the case and a non-significant result were obtained the interpretation would not be correct if one concluded there was no practically significant difference. The proper conclusion should be that there was insufficient power to detect practically important differences and that conclusive research would require a larger sample. If a significant result were obtained then, from a power viewpoint, the sample was large enough.

Example: In a Follow Through project 384 students were included in a comparison of Follow Through with non Follow Through students. This was the largest sample which could have been taken since there were only 384 children in the

program. Statistically significant results were obtained for the comparison of the Follow Through with non Follow Through students and the sample was therefore large enough with regard to power (had non significant results been obtained it would have been necessary to determine what effect sizes the test has a reasonable power for detect.)

8. Relation to a third variable -- In some cases it may not be possible to determine MES for the variable of interest but it may be possible to determine MES for another (third) variable and to determine the regression relationship between the dependent variable and the third variable. If this is the case it seems appropriate to determine what effect size in the currently investigated variable would predict a MES in the second variable; this effect size in the first variable would then be used as MES. The justification here would be that a MES change in the currently investigated variable would predict a MES change in the second variable. (The writer has never used this method.)

Example: The Nowicki-Strickland Internal-External Scale (IE) is relatively new with few studies done using the scale. Those studies which have been done have been largely correlational studies which would also yield regression equations. If one such study was between IE and IQ and the resulting equation were

$$IQ = 12 + 2\ IE$$

then a change of 2.5 units in IE would yield a change of 5 units in IQ. If 5 were the MES for IQ then 2.5 could be considered the MES for IE. A change of 2.5 in IE would then predict a change of 5 for IQ which is the MES for IQ.

9. Grade Equivalent -- In studies involving tests which have grade equivalent scores available and where the concern is with change, an upper bound to the MES would be the change in grade equivalent score which would normally take place during the time elapsed to conduct the study.

Example: A study of change in reading achievement used an achievement test which has grade equivalent scores available. The study covered a three month time period and it was decided to use a three month change in grade equivalent score as the MES.

10. Use 1 or 1/2.

Frequently the dependent variable is a test score. Test scores are usually counts of the number of correct answers. As counts, it is impossible to get a test score which differs by less than one point for two individuals. Thus if the results of the study are applicable to individuals, one could argue that the smallest meaningful effect is 1 unit. If one has a natural science background, it is perhaps more satisfying to use 1/2 as

the minimum effect size in these circumstances. Here the argument is that a test score is a rounded measurement and thus is only accurate to 1/2 of a unit. Both of these values for MES seem somewhat appealing as minimum effect sizes, even when considering the mean of a group rather than an individual. Imagine reporting a significant effect which represents a difference of 1 correct answer or less on a test (or imagine the difference to be less than 1/2 if that is your preference). Most researchers would find it difficult to defend anything smaller than this as a meaningful effect.

Example: Sample size is being determined for a study using the Nelson-Denny Reading test as the dependent variable. Sample size is determined using MES of 1/2.

11. Percent of one group above another -- Olejnik (1987) explained a method for determining a minimum effect size by deciding what percentage of subjects in an experimental group one might wish to exceed those in a control group. This might be easier stated by stating a proportion of the experimental group which exceeds the control group mean. Remember that is the two groups have the same means 50 % of the control group exceeds the control group mean (normal distributions and equal variances are assumed here). Thus if one wished that a particular treatment might result in 70% of the experimental group exceeding the mean of the control group, one has simply to determine the point to

locate the experimental group mean above the control group mean from a normal curve table. It is interesting to note that Olejnik's method does not require knowing either of the two means.

Example: A study is uses a dependent variable which has a standard deviation of 10. It is desired to determine a minimum effect size such that 80% of those in the experimental group have scores which exceed the control group. This will happen if when the difference between the two means is 8.4.

12. Select MES by using the mean effect size from a meta-analysis. If a meta-analysis has been done on the dependent variable of the study being designed, one might argue that the smallest significant effect from that meta-analysis represents a minimum effect size because it the expected effect.

Example: Matheny et al (1986) reported a meta-analysis of stress coping. The mean effect reported for groups administered treatment compared to control groups was .57. It should be noted that the mean effect size reported by Matheny is defined as the difference between the group means divided by a standard deviation. That measure would need to be rescaled by multipling the effect size by the standard deviation in order to be comparable with the other methods in this paper. This value, .57

times the standard deviation, is used as the MES for a study with stress coping as the dependent variable.

The following two methods for determining MES are last resorts which the writer feels should be used only if all other methods prove unfruitful. This is because in each case there is no justification for the results other than an arbitrary standard. However, all research must begin somewhere and if very little is known about the importance of differences in the dependent variable, it may be necessary to be quite arbitrary about the MES. In that case the following two methods seem as arbitrary as any. Both methods have the advantage of being easy to perform. The methods have advantages over the previous methods. The first is performable if a reliability measure is available and the last can always be done. Moreover, both methods make a clear statement to the reader of what power the researcher used in the study. This is perhaps the best justification of their use. (These are presented without example).

13. Standard Error of Measurement -- It the dependent variable is a test for which the standard error of measurement, SEM, has been estimated the SEM may be used as a MES. This is an arbitrary MES but perhaps may be justified by simply claiming it is as accurate as an individual can be measured on the dependent variable and therefore is a reasonable difference to seek. The

writer cannot justify using one SEM rather than two or 1.67, if just seems appealing to use one SEM and, if convenient, may make a nice convention.

14. Small, Medium, Large -- Cohen (1969, pp. 18-25) advocated the use of small, medium, and large effect sizes for the determination of power. He defines these as $.2\sigma$, $.5\sigma$, and $.8\sigma$ respectfully for testing differences in means using a t test (σ = common standard deviation of the two populations). Cohen has no justification for his definitions which appeals to the writer, his justification for his definitions appear convenient for making the tables which he makes but if the MES can be determined by one of the previous methods then it would seem unnecessary to be so arbitrary as to take a small, medium, or large effect size. If the reader were unable to justify a MES by any of the first 10 methods then he might select a small, medium, or large effect size a la Cohen and have a convenient description of what he had done for his readers. Moreover, the calculations would be simplified by using the tables which Cohen has prepared. Some light has been shed on small, medium,and large effect sizes by Brewer (1972) and comments in response to the Brewer article pointing out some of the fallacies of small, medium, and large effect sizes and of post hoc power calculations. (Cohen, 1973, Dayton, Schafer, and Rogers, 1973).

<u>What Size Power?</u>

Most researchers have found the probability of a type I error easy to handle; hypothesis testing has the choosing of a significance level build into the procedure. Selecting the significance level lets the researcher select a probabilistic protection against error. This has been easy to do because there is only one value of the tested parameter to consider: the hypothesized value. When considering significance level the question is only whether the hypothesized value is correct; the possible alternative values of the tested parameter are not considered when choosing a significance level. The possible alternative values only come under consideration when power or type II errors are examined.

The difficulty involved in examining power is that power differs for every possible alternative hypothesis. The researcher thus finds that the power of his test depends on what the true value of the tested parameter is. Of course if he knew the true value of the parameter it would not be necessary to do any hypothesis testing; therefore he finds himself in the position of speculating on the true value of parameter.

The previous section of this paper has listed 11 methods for choosing a minimal effect size. The advantage to choosing a minimal effect size is that the researcher can then concern himself with determining a sample size for that size alone and know that the test will have even greater power for a larger effect size. Determining power or sample size for the fixed

alternative of the minimum effect size is usually a sample matter and the calculation problem of calculating several powers is thus solved by calculating power only for the MES. The only question then is what size power to seek?

In selecting a significance level it has become somewhat traditional to select one of three levels: .05, .01, or .001. In most cases there has been no particular reason for selecting these probabilities of a type I error. They seem to have become the common values as a matter of tradition and that they represent levels of error which most researchers feel they can tolerate. Since most researchers can tolerate these levels for a type one error it seems to the writer that they could also tolerate these levels for a type II error. Probabilities for a type II error would thus be set at .05, .01, or .001. Since power is equal to 1 minus the probability of a type II error, power would then be set at .95, .99,or .999.

For some time educational researchers have lamented the fact that losses due to type I or type II errors are not usually quantifiable. Thus no relative importance has been able to be placed on type I and type II errors. This being the case, it would seem rather contradictory to set the probability of a type II error at any value other than that which we set the probability of a type I error. It is commonly accepted that research should be conducted with alpha at a maximum value of .05 (Cowles and Davis, 1982). I would say that all research should

also be conducted with beta at a maximum level of .05. This means that the minimum value of power should be .95 .

Whether or not this is an appropriate set of common values for power will be determined only when there is a body of research which has been conducted using these levels. It will be a matter of tradition and willingness to tolerate these error levels just as it has been with significance level. In a sense then we are currently establishing tradition, let us establish it with no greater probability of a type II error than what we have used for the type I error. Maximum tolerable probability for a type II error is thus .05 and minimum tolerable power is .95 .

Dangers of Arbitrary Selection of Minimal Effect Sizes

Some of this development of arguments may seem to be a lot of fuss for very little return. Why should anyone worry about making one of the arguments towards the front of the list? Why not simply use one of the last two methods? If a meta-analysis is not available use Cohen's small, medium, or large effect size and be done with the problem. The answer is that neither of these methods answers the question which is "What is the smallest difference that a researcher would wish to detect with a hypothesis test?" Without answering this question, the researcher is likely to make one of four errors

First, if the researcher selects too large a sample size, resources are likely to be wasted gathering

unnecessary data.

Second, if too large a sample size is selected, the researcher may conclude that a small difference has some practical significance when it does not. This would be much like one of those studies which makes much ado about a correlation of .09 which is significant at the .001 level.

Third, if the researcher selects too small a sample size, a difference which might be practically important be missed.

Fourth, if too small a sample size is selected and a difference is not detected, other researchers in the area may be discouraged from further work in the area and important areas of investigation may be omitted.

By selecting a minimal effect size and determining power for that effect, the researcher communicates to the reader what the researcher considers to be important and what the researcher's chances are for finding that effect or a larger one. Both the researcher and the scientific community are served well by this process. The researcher is assured of credibility whether statistically significant results are obtained or not. The scientific community benefits by having research which largely avoids the four errors listed above. The methods listed here can help the researcher make the argument as to what constitutes a minimum effect size.

References

Brewer, J. K. "On the power of statistical tests in the American Educational Research Journal" American Educational Research Journal 1972, 9, 391-401.

Brewer, J. K. & Owen, P. W. "A note on the power of statistical test in the Journal of Educational Measurement" Journal of Educational Measurement 1973, 10, 71-74.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences Academic Press, 1969.

Cohen, J. "Statistical power analysis and research results" American Educational Research Journal 1973, 10, 225-229.

Cowles, M. & Davis, C. "On the origins of the .05 level of statistical significance" American Psychologist, 1982, 37, 553-558.

Dayton, M. C. Schafer, W. D. & Rogers, B. G. "On appropriate uses and interpretations of power analysis: A Comment" American Educational Research Journal 1973, 10, 231-234.

Friedman, H. "Simplified determination of statistical power, magnitude of effect, and required sample sizes" Educational and Psychological Measurement 1982, 42, 521-528.

Hays, W. C. Statistics for the Social Sciences Holt, 1973.

Jensen, A. R. "How much can we boost I.Q. and scholastic achievement" Harvard Educational Review 1969, 39, 1-123.

Koele, P. "Calculating power in analysis of variance" Psychological Bulletin 1982, 92, 513-516

Kraemer, H. C. "A strategy to teach the concept and application of power of statistical tests" Journal of Educational Statistics 1985, 10, 173-195.

Matheney, K., Aycock, D.W., Pugh, J. L., Curlette, W. L., & Cannella, K. A. S. "Stress coping: a qualitative and quantitative synthesis with implications for treatment" The Counseling Psychologist 1986, 14, 499-549.

Neel, J. H. "Methods for Determining Minimal Effect Sizes in Power and Sample Size Calculations",

Olenjik, S. personal communication AERA 1987

Robey, R. R. & Barcikowski, R. S. "Calculating the Statistical Power of the Univariate and multivariate Repeated Measures Analysis of Variance" Educational and Psychological Measurement 1984, 44, 137-144.